心理科学进展 2021, Vol. 29, No. 1, 79–92
Advances in Psychological Science

https://dx.doi.org/10.3724/SP.J.1042.2021.00079

错误记忆产生的认知与神经机制：信息加工视角

郭 滢[1] 龚先旻[2] 王大华[1]

([1]北京师范大学发展心理研究院，北京 100875)
([2]香港中文大学何鸿燊海量数据决策分析研究中心&心理学系，香港 999077)

摘 要 采用信息加工视角，在划分不同信息来源的基础上分析编码、存储(巩固)、再激活/再巩固和提取的一系列加工过程如何导致错误记忆形成，由此总结出错误记忆产生的三个可能原因：(1)因缺乏针对目标事物特异性细节的记忆表征而侧重于编码和提取目标和非目标事物共享的抽象记忆表征，使被试更倾向于依赖抽象表征对缺失的目标细节进行重构，引发错误记忆；(2)目标事物启动了对应图式，导致与图式相关的非目标事物记忆表征得到增强，引发错误记忆；(3)误导信息干扰了再度激活状态下目标事物的记忆表征，妨碍其进行准确的记忆再巩固，从而引发错误记忆。未来研究可进一步探讨目标事物特异性细节的表征区域、不同类型的图式表征促进非目标事物记忆表征的具体机制以及提取阶段的图式复现对错误记忆形成的影响等问题。

关键词 错误记忆，认知图式，神经机制，信息加工

分类号 B842

1 引言

个体在回忆经历过的事件时，大脑所检索出的记忆往往并非是现实的忠实再现，而是掺杂了已有经验和外界干扰后对原始事件的重新建构。这可能导致个体错误地回忆起未经历过的事件或回忆的内容与现实并不相符，产生错误记忆。那么，错误记忆到底是怎么产生的呢？目前，已有部分学者从不同侧面对错误记忆产生的认知与神经机制进行了分析与梳理。如江荣焕和李晓东(2015)从模糊痕迹理论和关联激活理论出发，总结了关联性错误记忆的产生原因；王密和耿海燕(2010)则进一步分析了模糊痕迹理论中笼统的要点痕迹和精确的字面痕迹对关联性错误记忆的影响；刘振亮等人(2015)则梳理了导致植入性错误记忆产生的可能原因。这些综述无疑对我们理解错误记忆的产生机制有所裨益，但它们同样存在一些有待完善的地方：一方面，它们多围绕某一种错误记忆展开论述(如关联性错误记忆、植入性错误记忆等)，而缺乏涵盖多种错误记忆的整体框架(如：刘振亮 等, 2015)；另一方面，它们多侧重于刻画某个因素(如要点痕迹、字面痕迹等)对错误记忆的静态影响，分析“有或无”这个因素与错误记忆形成的关系，而缺乏对其动态作用机制的总结(如：王密，耿海燕, 2010)。

针对以上两点，本文采用信息加工视角：一方面，对不同的信息来源进行区分，分别论述对目标事件、内部图式和外界干扰的信息加工如何导致错误记忆形成。我们将由内部加工引发的错误记忆，如关联性错误记忆和由外部加工引发的错误记忆，如植入性错误记忆放在同一信息加工框架内进行探讨。这使本文不再像已有综述那样局限于具体的错误记忆种类。除此之外，不论是关联性错误记忆，植入性错误记忆，还是其它种类的错误记忆，都涉及对原始目标事件的信息加工。将这种正确记忆形成相关的信息加工与错误记忆形成相关的对内部图式和外界干扰的信息加工进行整合，有利于我们总结出更加全面系统的错误记忆产生机制。另一方面，对不同的信息加工过程进行区分，分别论述编码、存储、再激活/再巩固和提取的一系列加工过程中导致错误记忆的可能原因。记忆是个动态的过程，编码、存储、提取等任一环节出现差池都可能引发错误记忆。

收稿日期：2020-05-14
通信作者：王大华, E-mail: wangdahua@bnu.edu.cn

因此，在阐述错误记忆影响因素时不仅要考虑“有或无”，还要考虑“处于信息加工哪个阶段”，从而动态地分析其在错误记忆形成中的作用机制。

由此，本文将对信息来源的区分(见图 1 各灰色背景框)和对信息加工过程的区分(见图 1 各白色方框)相结合，在信息加工的框架内总结出三条错误记忆产生的可能机制(见图 1 各虚线方框)，每条机制着重描述某一来源信息的各加工过程。具体来说：机制 1 对应着对目标事件的加工，在编码、存储和提取阶段缺乏针对目标事件中特异性细节信息的记忆表征将导致错误记忆产生；机制 2 对应着对内部图式的加工，目标事件迅速启动图式后会在编码阶段改变其他相关事件的记忆表征，并通过在提取阶段再度激活图式来增加错误记忆；机制 3 对应着对外界干扰的加工，目标事件的记忆表征被再度激活时，对与目标事件不符的误导信息进行编码将干扰目标事件的原始记忆，进而引发错误记忆。下面我们将围绕这三条机制展开具体论述。

2 机制 1：缺乏针对目标事件的特异化细节表征

不同的记忆内容会在抽象的心理空间内形成不同的记忆表征，并与神经层面不同的神经表征相对应。本节所论述的特异化表征(distinctive representation)指针对某一特定记忆内容，能将之

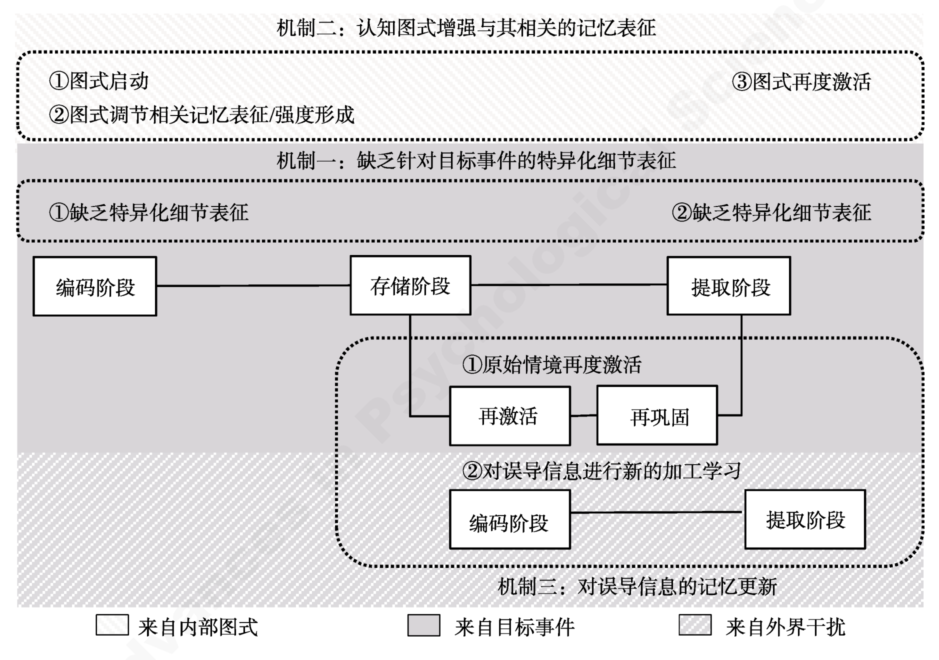


图 1 信息加工视角下错误记忆产生机制的示意图

注：各灰色背景框代表不同的信息来源(来自目标事件、内部图式和外界干扰)；各白色方框代表不同的信息加工过程(编码、存储、再激活/再巩固和提取)；各虚线方框代表错误记忆产生的可能机制，每条机制着重描述某一来源信息的各加工过程，由①②③代表，且它的发生阶段与下方的白色方框相对应。三条机制相互独立，又彼此联系 [1]，共同导致错误记忆的发生。有必要进行说明的是，机制 3 虽着重描述对外界干扰的信息加工，但同样涉及对目标事件的信息加工，因此代表机制 3 的虚线方框同时呈现在两种灰色背景框之上。

---

[1] 三种机制之间既相互独立，又彼此联系。其中，独立体现在：每条机制都对应着不同的信息来源，并与不同种类的错误记忆形成有关，比如机制 2 可以用来解释关联性错误记忆的形成，而机制 3 可以用来解释植入性错误记忆的形成。联系体现在：正确记忆形成相关的机制 1 与错误记忆形成相关的机制 2 和机制 3 通常共同发生。不论是关联性错误记忆，还是植入性错误记忆都缺乏针对目标事件的特异化细节表征。另外，机制 1 与机制 2、机制 1 与机制 3 间还存在相互影响的可能。如机制 2 中图式对特异性细节的抑制作用(如 van der Linden et al., 2017)以及机制 3 中误导信息与特异性细节间的相互竞争(如 Okado & Stark, 2005)。

区别于其他记忆内容的独特的记忆/神经表征。神经表征的特异化程度有高有低，特异化程度较低的神经表征仅能区别存在明显知觉/语义差异的记忆内容(如区别“悉尼歌剧院”和“中国长城”的图片)，而特异化程度较高的神经表征则能进一步区别存在某些细节差异的记忆内容(如区别形态相似但并不相同的“猫”的图片)。

错误记忆研究通常关注被试对相关诱饵(related lure，即与目标项目的知觉/语义特征高度相似但在某些细节上有所差别的探测，以下简称诱饵)的虚报/拒斥反应。拒斥诱饵，即区分诱饵和目标，依赖针对目标细节的特异化程度较高的神经表征。当存在这样的特异化表征时，目标与诱饵间记忆表征的匹配程度下降，被试对诱饵探测的记忆强度随之降低(Norman，2010)，进而更倾向于拒斥诱饵。反之，当缺乏这样的特异化表征而只存在反映目标和诱饵共同特征的抽象记忆表征时，被试便更倾向于依赖抽象表征对缺失的目标细节进行重构，导致虚报概率的增加。下面我们将分别论述编码阶段和提取阶段针对目标的特异化细节表征(以下简称特异化表征)在错误记忆产生中的内在机制及相关神经证据。

2.1　编码阶段：缺乏特异化表征

编码阶段缺乏特异化表征可能导致错误记忆形成。海马的相关研究为此提供了直接证据。记忆建构模型(constructive memory framework)强调海马的模式分离功能(pattern separation)在形成准确记忆中的重要作用(Schacter et al., 1998)。它是指海马将相似的记忆内容转换为正交化的神经表征的能力。具体来说，海马会尽可能地降低相似记忆内容间神经表征的重叠程度，从而使个体形成代表各记忆内容的标志性的特异化表征，这种具有高诊断性的特异化表征有助于不同记忆内容间的准确鉴别(LaRocque et al., 2013; Stevenson et al., 2020; Yassa et al., 2011)。Wing 等人(2020)的研究直接证实了上述机制的存在。在该研究中，被试首先学习若干语义类别下(如帐篷)的不同样例图片，一天后被试对学习过的目标图片、未学习过但语义类别与目标样例相同的诱饵图片以及未学习过且语义类别也不同于任何目标样例的无关图片进行再认。该研究使用多体素模式分析(multivoxel pattern analysis, MVPA)中的表征相似性分析[2](representational similarity analysis, RSA, Kriegeskorte et al., 2008)计算出编码阶段代表各目标样例间以及目标与诱饵样例间共同知觉/语义特征的类别表征，并发现当初级视皮层和顶上小叶表现出高水平的类别表征时，海马同样高水平的类别表征会使被试虚报诱饵，而海马低水平的类别表征(即海马的模式分离)则会使被试拒斥诱饵。由此可见，海马的模式分离能有效调节高度重叠的皮层表征是否会损害准确记忆的形成。但尽管如此，海马的模式分离也并非毫无限制。当记忆内容过于相似而超出分离极限时，海马就不得不转而启动模式完成功能(pattern completion)，复现相似表征并将之与当前表征进行比对，通过这种方式来使个体获得体现二者间细微差别的特异性细节信息，从而指导正确的记忆鉴别(van den Honert et al., 2016)。

除了海马，大量研究还发现一些皮层区域，如初级视皮层(或特定任务下的次级视皮层)也能对相似记忆内容的细节信息进行特异化编码(Baym & Gonsalves, 2010; Garoff-Eaton et al., 2005; Pidgeon & Morcom, 2016; St-Laurent et al., 2014; Xiao et al., 2017)。如 Baym 和 Gonsalves (2010)发现，次级视皮层能表征那些区别目标和诱饵的特异性细节信息(如目标事件为在某一故事背景下买“香蕉”，而诱饵事件为在相同背景下买“橘子”，“香蕉”即为目标事件区别于诱饵事件的特异性细节信息)。如果次级视皮层在编码阶段出现激活下降，那么就可能导致被试缺少对目标细节的特异化编码，从而增加对诱饵的虚报。

不过，也有不少研究并未发现初/次级视皮层在编码阶段的特异化表征对拒斥诱饵的预测作用(如 Pidgeon & Morcom, 2016)。一方面可能是因为编码阶段初级视皮层的特异化表征往往不能稳定地在提取阶段复现(Kuhl & Chun, 2014; St-Laurent

[2] 该方法将当前状态下多个体素的神经信号看作一个多维变量，即空间模式(spatial pattern)，并构建所有项目对间空间模式的表征差异性矩阵(representational dissimilarity matrix, RDM)。该矩阵还可来自其他状态下的的空间模式、项目属性模型、计算模型、行为模型等。该方法通过比较表征差异性矩阵间的关系，实现对当前状态下神经表征的解读(Kriegeskorte et al., 2008)。

et al., 2014; Staresina et al., 2012; Xiao et al., 2017); 另一方面可能是因为提取阶段初级视皮层的特异化表征往往不能直接影响记忆表现(Lee et al., 2019; Wing et al., 2020)。即便如此，我们也不能忽视初级视皮层在特异化记忆表征形成中的重要作用。它作为保留最多原始细节的加工区域，也可能通过影响海马的模式分离进而影响记忆表现(Gordon et al., 2014)。

2.2 提取阶段：缺乏特异化表征

提取阶段缺乏特异化表征也可能引发错误记忆，出现这种情况的原因可能有如下三种。(1)特异性细节的编码失败。大量研究表明，成功的记忆提取实际反映了编码阶段神经表征的复现(reactivation; Nyberg et al., 2000; Staresina et al., 2012; Wheeler et al., 2000)，因此编码阶段特异化表征的缺乏可能使提取阶段复现的神经表征同样过于抽象概括，在此基础上进行的记忆重构将增加被试对诱饵的虚报。Wing 等人(2020)使用表征相似性分析计算出编码阶段的类别表征(见 2.1)在提取阶段的复现程度，并发现海马处类别表征的神经复现越强，被试越倾向于虚报诱饵。这说明，错误记忆的产生伴随类别水平的抽象表征的神经复现，而在编码阶段形成具有高诊断性的特异化表征则是抑制错误记忆产生的必要非充分条件，被试能否在成功编码的基础上成功提取针对目标的特异性细节还受到存储质量和提取方式的影响。

(2)特异性细节的存储失败。研究表明，特异性细节信息随时间的衰减速度较快，程度较高，抗干扰性较差(Brainerd & Reyna, 1993, 2002; Sekeres et al., 2016)。因此，特异性细节信息可能在存储过程中出现丢失而无法提取，进而引发错误记忆。

(3)特异性细节的提取方式不当。还有研究表明，特异性细节信息并未完全丢失，它仍然存在于我们的记忆系统中，只是由于提取方式不当而无法通达(Gonsalves & Paller, 2000; Kensinger & Schacter, 2006; Slotnick & Schacter, 2004)。当给予被试合适的提取线索(如再次呈现目标)，目标项目的特异性细节依旧可以被顺利提取并抑制错误记忆的产生(Guerin et al., 2012a, 2012b; Sekeres et al., 2016; Weinstein et al., 2010; 陈红 等, 2015)。

以上三种原因虽然均可能导致提取阶段缺乏特异化表征，但不同的是，神经复现反映了提取阶段与编码阶段的直接联系，而存储缺失和提取不当则仅体现在提取阶段本身。下面我们将从神经层面具体论述在排除与编码阶段的直接联系(即海马的神经复现)后，提取阶段角回的神经表征如何影响错误记忆形成。不同于初级视皮层(见 2.1)，角回的功能发挥对记忆提取至关重要。一方面，Xiao 等人(2017)使用多体素模式分析中的表征连接性分析(representational connectivity analysis)发现编码阶段由视觉加工皮层负责采集的各目标项目间的内部表征结构在提取阶段转为由角回负责再现；另一方面，角回位于视觉加工皮层与负责认知调控的前额叶的交界处，因此更能对行为层面的记忆表现产生影响(Lee et al., 2019)。

角回在记忆提取中的重要作用敦促着研究者们去关注角回所包含的神经表征，因为这将直接影响记忆提取的质量。针对这一问题，已有大量研究发现提取阶段角回的神经表征的确具有特异性(Kuhl & Chun, 2014; Lee et al., 2019; Xiao et al., 2017)，但这些研究关注的特异化表征主要指能区别项目间明显知觉/语义差异(如“悉尼歌剧院”和“中国长城”的场景图片)的特异化表征，而非错误记忆研究关注的能区别项目间细节差异的特异化表征，二者在指代范畴上有所出入。

关于角回的特异化表征与错误记忆形成的具体关系，目前研究尚无定论。一方面，Ye 等人(2016)采用 Deese-Roediger-McDermott (DRM)范式在角回处发现，代表诱饵词汇从所有目标词汇处获得记忆痕迹的神经指标部分中介同一语义类别下所有词汇(包括目标和诱饵词汇)间的语义相似性对诱饵词汇记忆强度的预测关系。这可能说明，记忆提取时角回的特异化表征更多反映的是目标项目间明显的语义差异(Kuhl & Chun, 2014)，而这种程度较低的特异化表征并不能指导错误记忆研究中对诱饵探测的记忆鉴别。这与 Kurkela 和 Dennis (2016)通过元分析发现的角回在错误记忆提取中的一致性激活以及 McDermott 等人(2017)在 DRM 范式中发现的角回激活反映了对记忆探测的主观知晓感(perceived oldness)等研究结论一致。另一方面，也有研究认为角回的特异化表征体现了不同记忆内容间的细节差异(Richter et al., 2016)，并且被试可以利用这种特异性细节抑制对诱饵探测的虚报(Guerin et al., 2012a; Lee et al., 2019)。关于提取阶段角回的功能发挥对错误记忆的影响，目前研究尚未得出一致结论。未

来研究可尝试在错误记忆范式中考察角回的特异化表征以及与错误记忆形成的具体关系。

3　机制 2：认知图式增强与其相关的记忆表征

个体往往倾向于运用已有知识，即认知图式对信息进行自上而下的精细加工(elaborative processing)。这种图式驱动下的精细加工有助于个体对接收的信息进行快速有效的理解、吸收、组织和结构化，从而提升记忆效率，还有助于个体对事件主题和含义的提取，使抽象和概括成为可能。但另一方面这种精细加工也可能造成副产品，即错误记忆的产生(Schacter et al., 2011)。

图式指个体无数次经验累积所形成的体现事物间共同特征的高层级知识网络。它涵盖多种类型的已有知识，其中就包括错误记忆领域内备受关注的语义网络和脚本网络。前者体现概念间的联结关系及概念与类别间的从属关系，后者体现普通事件(如买杂货或订外卖)前后衔接、因果关联的一系列环节和环节中的典型场景(Gilboa & Marlatte, 2017)。

在神经生物层面，图式也指彼此联结的新皮层表征。不同类型的知识网络所涉及的新皮层表征区域也不尽相同。其中语义网络的关键中枢在额下回、颞中回/颞下回、颞极等，而脚本网络的关键中枢在内侧前额叶、后扣带回、压后皮质、楔前叶、顶下小叶等(Gilboa & Marlatte, 2017)。下面我们将具体论述图式导致错误记忆形成的内在机制及相关神经证据。

3.1　图式在错误记忆形成中的具体作用机制

3.1.1　编码阶段：图式启动

图式启动(schema instantiation)指激活认知图式并维持其作为通用的信息加工“模板”的过程(Ghosh et al., 2014)。图式启动在目标项目呈现后约 170 ms 时自发进行，它早于个体的意识觉察(Gilboa & Moscovitch, 2017)。腹内侧前额叶是图式启动的关键区域，它负责明确目标的对应图式(如词汇“预约”属于“看医生”，而非“睡觉”的图式，Ghosh et al., 2014)，识别目标与图式间的一致/不一致性(如“放在水池里的仙人掌”与仙人掌的图式不一致，Spalding et al., 2015)，并通过加强与颞中回/颞下回、顶下小叶、楔前叶、后扣带等后部皮层的功能连接强化与图式相关的信息加工(Bonasia et al., 2018; Gilboa & Moscovitch, 2017; Sommer, 2017)，弱化与图式无关或不符合图式的信息加工(Spalding et al., 2015; Sweegers et al., 2015; van der Linden et al., 2017)。当腹内侧前额叶受损而无法启动图式时，依赖图式驱动产生错误记忆的认知操作(如关联激活和/或要点抽取[3])将无法顺利进行。研究发现，在学习符合图式的目标图片后(如“放在烤箱里的披萨”)，腹内侧前额叶损伤患者反而会减少对相似诱饵图片的虚报(Spalding et al., 2015)，减少在 DRM 范式中对诱饵词汇的虚报和错误回忆(Warren et al., 2014)。应用重复经颅磁刺激(rTMS)对健康被试的腹内侧前额叶功能进行抑制也得到了相似的结果(Berkers et al., 2017)。

3.1.2　编码阶段：图式调节相关记忆表征/强度

图式作为信息加工的通用“模板”在启动后，一方面会自上而下地对信息加工过程进行在线指导，包括加强与图式相关的信息加工和减弱与图式无关的信息加工；另一方面还会进一步决定这样的加工过程将形成何种记忆表征/强度(本小节仅关注图式对记忆表征/强度的加强作用，减弱作用见 3.2)。图式指导在线加工为下一步调节记忆表征/强度的形成进行了加工时间和加工区域上的铺垫。

在加工时间上，图式指导在线加工于图式启动时(即目标呈现后约 170 ms)即已发生，而图式调节记忆表征/强度形成则发生在目标呈现后约 400 ms。此时，对符合图式的目标项目进行相关知识背景下的语义加工(semantic elaboration)将提高目标探测的记忆强度(Packard et al., 2017)。在加工区域上，从图式指导在线加工到图式调节记忆表征/强度形成发生了从腹内侧前额叶到后部皮层的转移(Gilboa & Marlatte, 2017)，腹内侧前额叶对后部皮层的神经活动起引领和捆绑的作用(Sommer, 2017; Baldassano et al., 2018)。

与图式启动相似，图式指导在线加工过程中腹内侧前额叶的功能发挥是图式调节记忆表征/

[3] 关联激活(associative activation)指个体在认知图式的指引下自发产生关联事物间激活扩散的过程；要点抽取(gist extraction)指个体在认知图式的驱动下主动抽取事物间共同特征的过程。二者均为图式驱动下可能导致虚报诱饵的认知操作(Gallo, 2006)。

强度的前提。其后，真正决定记忆表征/强度的是后部皮层的神经活动。并且，后部皮层的神经活动主要取决于目标的具体内容及进行的认知操作。具体来说，左侧额下回(Addis & McAndrews, 2006; Cooper & Ritchey, 2020; Sommer, 2017)和颞中回/颞下回(Binder & Desai, 2011)，尤其是颞极(Patterson et al., 2007)，主要负责利用语义网络形成目标词汇间的语义联系，而内侧前额叶、后扣带回、压后皮质、楔前叶、顶下小叶主要负责利用脚本网络还原目标事件的发生规律和典型场景(Baldassano et al., 2017; Baldassano et al., 2018; Chen et al., 2017)，两种图式网络分别决定各图式下记忆表征/强度的形成。下面我们将具体论述各区域对应的语义网络和脚本网络如何调节诱饵探测的记忆表征/强度，导致错误记忆的发生。

(1)语义网络对诱饵探测记忆表征/强度的调节。Kim 和 Cabeza (2007)要求被试学习一系列类别词表(如“农场动物”类别)并进行再认，结果发现，不论是击中目标还是虚报诱饵都会在编码阶段相同程度地激活负责语义加工的左侧额下回(Cabeza et al., 2001; Garoff-Eaton et al., 2007; Kubota et al., 2006)。这说明，类别词表启动的语义网络[4]不仅增强目标探测的记忆强度，还增强同一语义网络下诱饵探测的记忆强度。

除了左侧前额叶，研究还发现颞中回/颞下回(Dennis et al., 2008)，尤其是颞极在语义网络调节诱饵记忆表征中的重要作用(Chadwick et al., 2016; Zhu et al., 2019)。Chadwick 等人(2016)使用表征相似性分析在左侧颞极处发现，加工语义关联的目标词表与加工诱饵词汇间的神经表征相似性可以显著正向预测被试对该词表下诱饵词汇的虚报概率。这为颞极利用目标词表启动的图式表征促进相同图式下诱饵探测记忆表征的形成提供了具有说服力的证据。

(2)脚本网络对诱饵探测记忆表征/强度的调节。目前研究多从行为层面证明脚本网络对诱饵探测记忆强度的促进作用。具体来说，被试会根据脚本网络关联激活和/或要点抽取出符合脚本但实际未发生过的诱饵事件，产生错误记忆。如 Bower 等人(1979)发现当被试阅读一段去餐厅吃饭的文本后，已经对这一脚本了如指掌的被试会错误回忆出文本中并未提及的“点菜”环节(对脚本中前后衔接的环节形成错误记忆)；又如，Hannigan 和 Reinitz (2001)发现当被试见到描绘一女士从高高的橘子堆最底下抽橘子的场景后会虚报出见过橘子散落一地的场景(对脚本中因果关联的环节形成错误记忆)；Friedman (1979)发现当被试见到描绘某一特定场景(如儿童游戏室)的图片后会错误回忆出该场景中未呈现过的典型物品(如泰迪熊，脚本中典型场景引发的错误记忆)。

据我们所知，目前只有一项研究从神经层面对基于脚本的错误记忆形成进行了探讨。Aminoff 等人(2008)首先为被试呈现一系列成对出现的物品图片，要求被试为成对图片构建一个共同场景，并根据场景典型与否进行强场景框架(如“推土机”与“警示柱”)和弱场景框架(如“相机”与“剪刀”)的划分。一天后，被试对强/弱场景框架的目标图片(如推土机或相机)、与强场景框架相关但未呈现过的诱饵图片(如警示牌)以及与强/弱场景框架图片无关也未呈现过的无关图片(如吊灯)进行再认。结果表明，相比无关图片，被试会更多虚报见过诱饵图片，并且这种错误记忆的产生区域(虚报 vs. 正确拒斥)与典型场景的加工区域(强 vs.弱场景框架)存在重叠，二者均激活内侧前额叶、压后皮质及顶下小叶。因此，该研究间接证明了典型场景启动的脚本网络会促进相同网络下诱饵图片的记

[4] 关键区域，如额下回、颞中回/颞下回及内侧前额叶、压后皮质、顶下小叶的激活能否代表相应图式在发挥作用：不论从心理表征还是从神经表征层面，成为图式的一个必备要素即为具有关联性的网络组织结构(Ghosh & Gilboa, 2014)，因此即使是关键区域也不能简单代表相应图式本身，但这并不意味着不能代表相应图式在发挥作用，原因有二：(1)关键区域的激活由图式驱动。以左侧额下回为例，研究发现，当关联加工的形成负荷越大(如形成玩具、百合花、羊毛间的语义联系并进行关联记忆)，左侧额下回的激活就越强(Addis & McAndrews, 2006)。尽管在神经结果上只观察到左侧额下回的激活，但其激活实则反映了语义网络的推动作用；(2)关键区域的激活是相对的，体现了图式的相对强弱。以压后皮质等区域为例，研究发现，关联记忆典型场景内的两个物品(如“推土机”与“警示柱”)比关联记忆非典型场景内的两个物品(如“相机”与“剪刀”)更多激活压后皮质等相关区域(Aminoff et al., 2008)。尽管在神经结果上同样只观察到压后皮质等区域的激活，但其激活实则体现了经验累积所形成的脚本网络的相对强弱。因此，尽管关键区域不等同于相应图式本身，却能代表相应图式在发挥作用。

忆强度。不过, 该研究并未像 Chadwick 等人(2016)那样直接探讨图式表征与相同图式下诱饵探测记忆表征间的关系。未来研究可进一步使用多体素模式分析探讨跨模态(Baldassano et al., 2017; Baldassano et al., 2018)、跨阶段(Baldassano et al., 2017; Chen et al., 2017)、跨被试(Baldassano et al., 2018; Chen et al., 2017)的脚本表征如何促进诱饵探测记忆表征的形成。

3.1.3 提取阶段: 图式再度激活

研究发现, 当被试再认符合图式的目标图片(van der Linden et al., 2017)或回忆符合图式的目标事件时(Chen et al., 2017), 图式会在提取阶段再度激活并通过加强与后部皮层的功能连接促进对目标探测的击中(Bonasia et al., 2018; van Kesteren et al., 2010)。这种现象同样会发生在再认诱饵探测时。如提取阶段语义网络的关键区域, 即左侧前额叶及颞中回/颞下回的再度激活会促进对诱饵探测的虚报(Garoff-Eaton et al., 2007; Moritz et al., 2006; Webb et al., 2016)。但据我们所知, 目前尚未有研究使用多体素模式分析直接探讨提取阶段神经复现的图式表征与相同图式下诱饵探测记忆表征间的关系。

与上述研究思路不同, van Buuren 等人(2014)发现空间图式的存在会使被试更慢但更准确地提取图式相关探测, 并且这种高度策略化的提取过程会激活负责自上而下注意调控的顶上小叶和负责认知监控与评价的背外侧前额叶。在错误记忆领域, 上述区域的神经活动常见于对诱饵探测的回想拒斥(Bowman & Dennis, 2016), 这是否说明提取阶段图式的再度激活也可能通过加强监控而抑制错误记忆产生? 未来研究可进一步关注图式在提取阶段的再度激活对错误记忆的影响。

3.2 图式与特异性细节的关系

在 3.1.2 中我们具体论述了图式相关记忆表征/强度的加强过程。在本节中, 我们将转而论述图式无关记忆表征/强度的减弱过程以及该过程导致错误记忆产生的可能原因(Spalding et al., 2015)。错误记忆研究中区别目标与诱饵的特异性细节通常无益于相应图式的识别, 因此会作为与图式无关的信息受到加工抑制。研究发现, 在图式启动的过程中, 与图式加工相关的角回会抑制海马的神经活动, 使海马逐渐退出对目标图片的编码过程, 并且角回对海马的抑制作用越强, 被试越倾向于虚报见过相同图式下未呈现过的诱饵图片(van der Linden et al., 2017)。Doss 等人(2018)认为这是由于图式加工流畅性的提升干扰了海马的模式分离, 使海马无法形成针对细节的特异化表征, 进而增加错误记忆(见 2.1)。但这并不意味着缺乏特异化表征仅作为图式调节记忆表征的副产品影响错误记忆产生。当目标项目的呈现方式不足以启动相应图式时(如间隔呈现两张相似图片), 特异化表征也可以独立影响错误记忆形成(van den Honert et al., 2016)。

4 机制 3: 对误导信息的记忆更新

由于个体时刻处于与外界环境的交互中, 个体形成的记忆不仅会受内部图式的影响, 还会受外界环境变化的影响。面对环境变化, 动态的记忆系统会适应性地进行记忆更新, 及时将环境中新的有用信息灵活地纳入已有记忆表征中, 从而更好地指导行为实践。但有时环境中的新信息是与原始事件, 即目标事件不符的误导信息, 这时原本具有适应性的记忆更新过程就可能由于纳入了误导信息而造成副产品, 即错误记忆的产生(Schacter et al., 2011)。

对这种错误记忆的考察通常使用误导信息干扰范式(原始事件—信息误导—记忆测试)。具体来说, 在完成对原始事件的学习后, 被试会再次学习改动后包含误导信息的原始事件或在包含误导信息的诱导性提问下回想原始事件的相关内容, 从而内在地诱发原始事件发生情境的再度激活(见 4.2.1), 并促使被试对与原始事件不符的误导信息进行新的加工学习(见 4.2.2)。这将导致被试在针对原始事件的记忆测试中由于纳入了误导信息而发生误导信息对原始记忆的闯入(intrusion), 表现出对误导信息的虚报。在对这一过程展开详细论述之前, 我们有必要明确几个记忆再巩固理论下发生记忆更新必须满足的边界条件。

4.1 记忆更新必须满足的边界条件

记忆再巩固理论(memory reconsolidation theory)认为已巩固的记忆再度激活后将在数分钟内发生蛋白质分解, 此时记忆会暂时重返不稳定状态, 只有经过数小时新蛋白质的合成, 记忆才会恢复原来的巩固状态(Hardt et al., 2010)。在原

始记忆再度激活而变得脆弱的时间窗口内，原始记忆可能发生记忆强度上的动态改变，如记忆增强(Jonker et al., 2018; Koen & Rugg, 2016; Kuhl et al., 2010)和记忆减弱(Kim et al., 2014)，也可能发生记忆内容上的动态改变，如记忆更新(如 Sinclair & Barense, 2018)。

其中，对误导信息进行记忆更新必须满足三个边界条件：(1)原始记忆的再度激活。根据记忆再巩固理论，只有在原始记忆再度激活进入不稳定状态的时间窗口内，原始记忆才可能发生包括记忆更新在内的一系列动态改变。(2)存在超出原始记忆内容的误导信息。研究发现，当原始记忆受到超出其记忆内容的其它信息的干扰，并对干扰信息进行新的加工学习时，再度激活的原始记忆更倾向于发生记忆内容，而非记忆强度的改变(Hardt et al., 2010)。在误导信息干扰范式中，被试对与原始事件不符的误导信息进行新的加工学习将为原始记忆的内容改变提供必要的“素材”。(3)原始记忆的再度激活强度处于中等水平。记忆的非单调可塑性理论(nonmonotonic plasticity hypothesis, NMPH)认为记忆的再度激活强度与其发生的动态改变之间存在 U 型非线性关系，再度激活较弱时不会引发记忆改变，再度激活较强时反而会减弱记忆，而只有再度激活很强时才会增强记忆(Ritvo et al., 2019)。在误导信息干扰范式中，被试仅学习过一次原始事件，因此原始记忆的再度激活强度通常会落入“较弱—较强”的区间内(Kim et al., 2014)。在这种情况下，原始记忆更倾向于发生记忆减弱，并在此基础上受误导信息干扰，发生记忆更新。

本文采用信息加工视角，关注对不同来源信息进行编码、存储、再激活/再巩固和提取的动态过程，为避免内容分散，本文不再针对再度激活强度展开论述，但这仍是未来研究中的一个重要课题。下面我们将具体论述原始情境的再度激活以及对误导信息的加工学习如何导致错误记忆形成，其中前者是错误记忆形成的时间窗口，后者则为错误记忆形成提供内容素材，二者缺一不可。

4.2 更新误导信息形成错误记忆的具体机制

4.2.1 时间窗口：原始情境再度激活

误导信息干扰范式会要求被试再次学习改动后的原始事件或通过提问促使被试回想原始事件的相关内容，从而内在地诱发原始情境的再度激活[5]。作为范式涉及的固有认知过程，原始情境的再度激活很难与范式本身拆解开来，这一定程度上阻碍了我们进一步探讨该过程与错误记忆形成(误导信息对原始记忆的更新/闯入)的具体关系。下面我们介绍的范式则可以较好地解决这一问题，并进一步证实原始情境再度激活是发生记忆闯入的关键窗口，这一方面体现在该窗口的开放与否至关重要，另一方面则体现在该窗口何时开放至关重要。

首先，该窗口的开放与否，即原始情境再度激活与否至关重要。在 Hupbach 等人(2009)的研究中[6]，所有被试首先学习一组物品(组 1)，两天后提示组被试在相同实验环境下学习另一组物品(组 2)，并在学习前回想两天前是如何学习组 1 的，无提示组被试则在不同实验环境下直接开始组 2 的学习。又过两天后所有被试进行再认测试，并

---

[5] 本节关注的原始情境再度激活是对记忆再巩固理论中关于原始记忆再度激活的一种细化，这种细化尤其适用于误导信息干扰范式，这是因为这种范式涉及的原始事件通常内容极为丰富(如呈现一系列描绘某一日常活动的图片)，任何形式下的再度激活都无法完美再现原始事件的全部内容(Sinclair & Barense, 2019)，而更多体现为原始事件发生情境的再现(Jacques et al., 2013)。相比之下，记忆再巩固框架下的其他经典范式，如“AB-BC”范式涉及的原始事件通常较为简单(如成对图片)，因此更适合探讨原始事件内容本身再度激活对记忆动态改变的影响。

[6] 除了记忆再巩固理论，时序背景模型(temporal context modal, TCM)同样可以为 Hupbach 等人(2009)的研究结果提供合理解释。具体来说，对两组物品的学习不仅会形成针对物品本身的记忆，还会形成物品与物品呈现的时序背景间的联结记忆。提示组在学习组 2 物品之前复现组 1 物品呈现的时序背景，并在这一时间窗口内对组 2 物品进行新的加工学习，组 1 背景与组 2 物品发生捆绑，导致组 1 物品和组 2 物品由于共享相同的组 1 背景而相通，进而增加了组 2 物品闯入组 1 记忆的可能。并且，组 1 和组 2 物品间学习顺序的关系使组 2 物品可以与组 1 背景发生捆绑，组 1 物品却不可以与组 2 背景发生捆绑，因此提示组只单方向地表现出组 2 对组 1 记忆的闯入(Sederberg et al., 2011)。时序背景模型对于背景复现以及对新项目进行加工学习的关注与记忆再巩固理论基本一致，但不同的是，时序背景模型并不假定新信息对原始记忆进行了修改与更新，而认为二者只是利用赫布突触效应加强联结，原始记忆内容保持相对完好。并且，时序背景模型不像基于蛋白质合成与分解的记忆再巩固理论那样强调记忆更新的时间依赖性。Sinclair 和 Barense (2018)发现两种理论模型可以共同解释记忆闯入的发生。

进一步回忆旧物品来自组 1 还是组 2。该研究发现，相比无提示组，提示组会更多将组 2 物品虚报为来自组1，但不会更多将组1物品虚报为来自组 2。提示组之所以单向地表现出组 2 闯入组 1 记忆是由于学习组 2 时组 1 学习情境的再度激活开放了组 2 干扰组 1 记忆的时间窗口，从而使记忆闯入的发生成为可能。

其次，该窗口何时开放，即原始情境在误导信息呈现之前还是之后再度激活同样至关重要。研究发现，只有原始情境在误导信息呈现之前再度激活(Sinclair & Barense, 2018)，并延续至误导信息呈现阶段才会增加误导信息闯入原始记忆的概率(Jacques et al., 2013)。Gershman 等人(2013)使用多体素模式分类[7](multivoxel pattern classification, 雷威 等，2010)进一步将原始情境神经复现的准确时间锁定在误导信息呈现前 2 秒。该研究沿用 Hupbach 等人(2009)的范式发现，组 2 物品呈现 2 秒前组 1 学习情境的神经复现程度越高，该组 2 物品就越容易闯入对组 1 物品的记忆，并且这种神经复现对记忆闯入的预测作用仅发生在组 2 物品呈现之前的 2 秒，组 2 物品呈现之时及之后的神经复现都不能使该组 2 物品闯入组 1 记忆。

4.2.2 内容素材：对误导信息进行新的加工学习

Lee (2009)指出记忆再巩固并非简单指代记忆从再度激活到自动恢复稳态的过程，而是在为记忆不断修改更新提供条件，因此只有当存在超出原始记忆内容的干扰信息时才会发生记忆更新，对这些新信息进行加工学习将为原始记忆的内容改变提供必要素材。误导信息干扰范式中的误导信息与原始事件不符，因此也属于记忆更新必需的新信息。在编码这些信息时投入更多认知资源，进行更多的认知/神经加工将促进对误导信息的虚报，产生错误记忆。

Okado 和 Stark (2005)使用误导信息干扰范式发现，如果被试在编码原始事件时更多激活左侧海马体尾部及嗅周皮质，就更有可能拒斥诱饵事件，而如果被试在编码包含误导信息的诱饵事件时更多激活左侧海马体尾部及嗅周皮质，就更有可能虚报诱饵事件，产生错误记忆。这种交互作用的存在说明左侧海马尾部及嗅周皮质负责加工区别原始事件和诱饵事件的特异性细节信息(如小偷偷了女孩钱包并躲在“门”后还是“树”后)，拒斥诱饵事件是由于编码原始事件时的特异化细节表征(见 2.1)，而虚报诱饵事件是由于编码诱饵事件中的误导细节时进行了过多认知/神经加工。而如果编码诱饵事件时能更多激活默认网络，如前、后扣带及左侧楔前叶等内侧皮层，从而脱离当前对误导信息的认知/神经加工，被试就更有可能拒斥诱饵事件中的误导信息，减少错误记忆(Baym & Gonsalves, 2010)。

5 结论与反思

采用信息加工视角，本文详细论述了对不同来源信息(来自目标事件、内部图式和外界干扰)进行编码、存储、再激活/再巩固和提取的一系列加工过程中导致错误记忆产生的可能原因。这一方面使本文在论述错误记忆的认知机制时，比前人综述更加强调信息加工的动态性且较少受限于具体的错误记忆种类(如：刘振亮 等，2015)。另一方面，本文在论述错误记忆的神经机制时，并未从错误记忆与正确记忆的神经活动异同这一传统方向入手，以描绘错误记忆的“神经画像” (如 Dennis et al., 2015)，而旨在揭示错误记忆的产生原因。因此，我们从对信息的认知加工视角切入，并始终在信息加工框架内进行论述，这使我们可以更加全面系统地展现错误记忆的产生原因。不仅如此，据我们所知，前人在综述错误记忆神经机制时几乎无一例外地仅涵盖使用单体素激活分析(univariate activation)的相关研究(Dennis et al., 2015; Johnson et al., 2012; Straube, 2012)。受方法学的限制，这类研究虽然可以较好地揭示诱饵探测记忆强度的形成原因，但对其具体内容的揭示却十分有限(Davis & Poldrack, 2013)。若想针对记忆内容进行研究，则有必要使用多体素模式分析挖掘相应神经表征中蕴含的信息。本文引用了大量使用这种分析方法的前沿研究，使我们可以进一步从内容层面揭示诱饵探测记忆表征的形成原因。这也是对前人综述一个较好的补充和拓展。

---

[7] 该方法借鉴机器学习研究中的模式分类(pattern classification)技术，利用不同认知状态下的空间模式训练分类器(classifier)，再用独立的实验数据测试分类器的性能。这种方法不仅对认知表征差异的检测力度更强，而且可以使研究者根据分类器提供的证据反过来从神经信号中推测被试的认知状态，有利于我们深入理解认知状态在大脑中的表征(雷威 等, 2010)

总结全文，我们认为错误记忆的产生可能出于以下三个原因：其一，海马未能在编码阶段形成并在提取阶段复现针对目标的特异化细节表征，而只存在反映目标和诱饵共同特征的抽象记忆表征，使被试更倾向于依赖后者对缺失的细节信息进行重构，增加错误记忆。另外，提取阶段角回的神经表征与错误记忆形成的具体关系是当前领域内的热点话题之一。未来研究可尝试在错误记忆范式中考察角回的特异化表征与错误记忆的关系。并且，由于角回的神经表征受任务要求的调节(Kuhl et al., 2013)，未来研究还可关注角回与前额叶的功能交互对错误记忆形成的影响。另外，研究者们还可进一步思考这样一个问题：存在针对细节的特异化表征是否必然导致错误记忆减少？如视觉呈现的各目标词汇的特异性细节能否指导听觉呈现的目标与诱饵词汇间的记忆鉴别？

其二，目标项目启动图式后，在腹内侧前额叶的引领和捆绑下，语义网络(额下回、颞中回/颞下回、颞极等)和脚本网络(内侧前额叶、后扣带回、压后皮质、楔前叶、顶下小叶等)使与图式相关的诱饵探测记忆表征被增强，与图式无关的目标项目细节表征被减弱，导致错误记忆产生。目前该方向的研究趋势为：(1)相较于脚本网络，更关注语义网络驱动下错误记忆的形成；(2)相较于提取阶段图式复现对错误记忆的影响，更关注编码阶段图式对诱饵探测记忆表征的调节作用。未来研究可使用多体素模式分析探讨脚本表征如何促进诱饵探测记忆表征的形成，以及提取阶段神经复现的图式表征是通过促进相同图式下诱饵探测的记忆表征，进而增加错误记忆，还是会通过加强对诱饵探测的监控而抑制错误记忆产生？另外，研究者们还可进一步斟酌这样一个问题：图式是否必然增加与图式吻合的错误记忆？Hannigan 和 Reinitz (2001)发现被试存在事件发生因果规律的脚本，在呈现事件发生的原因后，他们会错误回忆出事件产生的结果。但同时脚本的存在也会使被试在呈现结果后形成对原因的预期(实际并未呈现原因)，从而减少对原因的虚报。图式导致预测误差(prediction error)的形成是否会抑制错误记忆形成？负责自动监测预期和真实事件间匹配程度(match/mismatch detection)的海马 CA1 分区在其中发挥怎样的作用(Chen et al., 2011)？

其三，在原始情境再度激活的时间窗口内对误导信息进行新的加工学习，干扰记忆再巩固过程，促使被试将误导信息更新至原始记忆中，产生错误记忆。目前该方向的研究较少关注误导信息的提取过程。已有部分研究在提取阶段发现记忆表征间存在相互竞争，并且相对表征强度能直接预测被试的行为报告(Kuhl et al., 2012; Kuhl et al., 2011)。延续这一思路，未来研究可尝试探讨提取阶段是否存在误导信息与原始信息间的竞争以及这种竞争如何影响对误导信息的虚报。另外，尽管并非本文讨论的重点，对原始记忆再度激活强度的探讨也许可以提示我们减少错误记忆的可行方法，如根据记忆的非单调可塑性理论(见 4.1)，过度学习原始事件使原始记忆的再度激活达到可以发生记忆增强的范围。这也许可以帮助抑制对误导信息的记忆更新。除此之外，Bridge 和 Voss (2014)发现只有在主动提取时占主导地位的记忆才会与新信息结合。在误导信息干扰范式中，被试主动回想原始事件的相关内容时通常会优先提取出事件的发生背景，而非具体细节(Putnam et al., 2017)。此时，相比细节记忆，占主导地位的背景记忆与误导信息间的结合是否是错误记忆形成的深层原因？对这一问题的探究有助于深化我们对于误导信息更新与错误记忆形成的理解。

致谢：感谢 Anastasia Besika 对英文摘要进行的细致校阅。

参考文献

陈红，郭春彦，杨海波．(2015). 延迟间隔和提取条件对短时错误记忆的影响．*心理与行为研究*, *13*(1), 37–43.

江荣焕，李晓东．(2015). 错误记忆的发展性逆转：为什么越长大越易“错”？*心理科学进展*, *23*(8), 1371–1379.

雷威，杨志，詹旻野，李红，翁旭初．(2010). 利用脑成像多体素模式分析解码认知的神经表征：原理和应用．*心理科学进展*, *18*(12), 1934–1941.

刘振亮，刘田田，韩佳慧，沐守宽．(2015). 错误记忆的可植入性．*心理科学进展*, *23*(5), 806–814.

王密，耿海燕．(2010). 从关联性记忆错觉的毕生发展看记忆的适应性特质．*科学通报*, *55*(4), 307–315.

Addis, D. R., & McAndrews, M. P. (2006). Prefrontal and hippocampal contributions to the generation and binding of semantic associations during successful encoding. *Neuroimage*, *33*(4), 1194–1206.

Aminoff, E., Schacter, D. L., & Bar, M. (2008). The cortical underpinnings of context-based memory distortion. *Journal of Cognitive Neuroscience*, *20*(12), 2226–2237.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron, 95*(3), 709–721.

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *Journal of Neuroscience, 38*(45), 9689–9699.

Baym, C. L., & Gonsalves, B. D. (2010). Comparison of neural activity that leads to true memories, false memories, and forgetting: An fMRI study of the misinformation effect. *Cognitive, Affective, & Behavioral Neuroscience, 10*(3), 339–348.

Berkers, R. M. W. J., van der Linden, M., de Almeida, R. F., Müller, N. C. J., Bovy, L., Dresler, M., ... Fernández, G. (2017). Transient medial prefrontal perturbation reduces false memory formation. *Cortex, 88*, 42–52.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536.

Bonasia, K., Sekeres, M. J., Gilboa, A., Grady, C. L., Winocur, G., & Moscovitch, M. (2018). Prior knowledge modulates the neural substrates of encoding and retrieving naturalistic events at short and long delays. *Neurobiology of Learning and Memory, 153*, 26–39.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*(2), 177–220.

Bowman, C. R., & Dennis, N. A. (2016). The neural basis of recollection rejection: Increases in hippocampal-prefrontal connectivity in the absence of a shared recall-to-reject and target recollection network. *Journal of cognitive neuroscience, 28*(8), 1194–1209.

Brainerd, C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review, 100*(1), 42–67.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*(5), 164–169.

Bridge, D. J., & Voss, J. L. (2014). Hippocampal binding of novel information with dominant memory traces can support both memory stability and change. *Journal of Neuroscience, 34*(6), 2203–2213.

Cabeza, R., Rao, S. M., Wagner, A. D., Mayer, A. R., & Schacter, D. L. (2001). Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proceedings of the National Academy of Sciences of the United States of America, 98*(8), 4805–4810.

Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences of the United States of America, 113*(36), 10180–10185.

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience, 20*(1), 115–125.

Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. *Learning & Memory, 18*(8), 523–528.

Cooper, R. A., & Ritchey, M. (2020). Progression from feature-specific brain activity to hippocampal binding during episodic encoding. *Journal of Neuroscience, 40*(8), 1701–1709.

Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: practices and pitfalls. *Annals of the New York Academy of Sciences, 1296*(1), 108–134.

Dennis, N. A., Bowman, C. R., & Turney, I. C. (2015). Functional neuroimaging of false memories. In D. R. Addis, M. Barense & A. Duarte (Eds.), *The Wiley Handbook on the Cognitive Neuroscience of Memory* (pp. 150–171). Hoboken, NJ: John Wiley & Sons, Ltd.

Dennis, N. A., Kim, H., & Cabeza, R. (2008). Age-related differences in brain activity during true and false memory retrieval. *Journal of cognitive neuroscience, 20*(8), 1390–1402.

Doss, M. K., Picart, J. K., & Gallo, D. A. (2018). The dark side of context: Context reinstatement can distort memory. *Psychological Science, 29*(6), 914–925.

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*(3), 316–355.

Gallo, D. A. (2006). Processes that cause false memory. In H. L. Roediger & J. R. Pomerantz (Eds.), *Associative illusions of memory: False memory research in DRM and related tasks* (pp. 39–73). New York, NY: Psychology Press.

Garoff-Eaton, R. J., Kensinger, E. A., & Schacter, D. L. (2007). The neural correlates of conceptual and perceptual false recognition. *Learning & Memory, 14*(10), 684–692.

Garoff-Eaton, R. J., Slotnick, S. D., & Schacter, D. L. (2005). The neural origins of specific and general memory: The role of the fusiform cortex. *Neuropsychologia, 43*(6), 847–859.

Gershman, S. J., Schapiro, A. C., Hupbach, A., & Norman, K. A. (2013). Neural context reinstatement predicts memory misattribution. *Journal of Neuroscience, 33*(20), 8590–8595.

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia, 53*, 104–114.

Ghosh, V. E., Moscovitch, M., Colella, B. M., & Gilboa, A. (2014). Schema representation in patients with ventromedial

PFC lesions. *Journal of Neuroscience, 34*(36), 12057–12070.
Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences, 21*(8), 618–631.
Gilboa, A., & Moscovitch, M. (2017). Ventromedial prefrontal cortex generates pre-stimulus theta coherence desynchronization: A schema instantiation hypothesis. *Cortex, 87*, 16–30.
Gonsalves, B., & Paller, K. A. (2000). Neural events that underlie remembering something that never happened. *Nature Neuroscience, 3*(12), 1316–1321.
Gordon, A. M., Rissman, J., Kiani, R., & Wagner, A. D. (2014). Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cerebral Cortex, 24*(12), 3350–3364.
Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012a). Interactions between visual attention and episodic retrieval: dissociable contributions of parietal regions during gist-based false recognition. *Neuron, 75*(6), 1122–1134.
Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012b). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language, 66*(1), 68–78.
Hannigan, S. L., & Reinitz, M. T. (2001). A demonstration and comparison of two types of inference-based memory errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(4), 931–940.
Hardt, O., Einarsson, E. Ö., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology, 61*(1), 141–167.
Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: Updating or source confusion? *Memory, 17*(5), 502–510.
Jacques, P. L. S., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences of the United States of America, 110*(49), 19671–19678.
Johnson, M. K., Raye, C. L., Mitchell, K. J., & Ankudowich, E. (2012). The cognitive neuroscience of true and false memories. In R. F. Belli (Ed.). *True and false recovered memories: Toward a reconsolidation of the debate* (pp. 15–52). New York, NY: Springer.
Jonker, T. R., Dimsdale-zucker, H., Ritchey, M., Clarke, A., & Ranganath, C. (2018). Neural reactivation in parietal cortex enhances memory for episodically linked information. *Proceedings of the National Academy of Sciences of the United States of America, 115*(43), 11084–11089.
Kensinger, E. A., & Schacter, D. L. (2006). Neural processes underlying memory attribution on a reality-monitoring task. *Cerebral Cortex, 16*(8), 1126–1133.
Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences of the United States of America, 111*(24), 8997–9002.
Kim, H., & Cabeza, R. (2007). Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cerebral Cortex, 17*(9), 2143–2150.
Koen, J. D., & Rugg, M. D. (2016). Memory reactivation predicts resistance to retroactive interference: evidence from multivariate classification and pattern similarity analyses. *Journal of Neuroscience, 36*(15), 4389–4399.
Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2*, 4.
Kubota, Y., Toichi, M., Shimizu, M., Mason, R. A., Findling, R. L., Yamamoto, K., & Calabrese, J. R. (2006). Prefrontal hemodynamic activity predicts false memory—A near-infrared spectroscopy study. *Neuroimage, 31*(4), 1783–1789.
Kuhl, B. A., Bainbridge, W. A., & Chun, M. M. (2012). Neural reactivation reveals mechanisms for updating memory. *Journal of Neuroscience, 32*(10), 3453–3461.
Kuhl, B. A., & Chun, M. M. (2014). Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *Journal of Neuroscience, 34*(23), 8051–8060.
Kuhl, B. A., Johnson, M. K., & Chun, M. M. (2013). Dissociable neural mechanisms for goal-directed versus incidental memory reactivation. *Journal of Neuroscience, 33*(41), 16099–16109.
Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences of the United States of America, 108*(14), 5903–5908.
Kuhl, B. A., Shah, A. T., DuBrow, S., & Wagner, A. D. (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nature neuroscience, 13*(4), 501–506.
Kurkela, K. A., & Dennis, N. A. (2016). Event-related fMRI studies of false memory: An Activation Likelihood Estimation meta-analysis. *Neuropsychologia, 81*, 149–167.
LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *Journal of Neuroscience, 33*(13), 5466–5474.
Lee, H. M., Samide, R., Richter, F. R., & Kuhl, B. A. (2019).

Decomposing parietal memory reactivation to predict consequences of remembering. *Cerebral Cortex, 29*(8), 3305–3318.

Lee, J. L. C. (2009). Reconsolidation: maintaining memory relevance. *Trends in Neurosciences, 32*(8), 413–420.

McDermott, K. B., Gilmore, A. W., Nelson, S. M., Watson, J. M., & Ojemann, J. G. (2017). The parietal memory network activates similarly for true and associative false recognition elicited via the DRM procedure. *Cortex, 87*, 96–107.

Moritz, S., Gläscher, J., Sommer, T., Büchel, C., & Braus, D. F. (2006). Neural correlates of memory confidence. *Neuroimage, 33*(4), 1188–1193.

Norman, K. A. (2010). How hippocampus and cortex contribute to recognition memory: revisiting the complementary learning systems model. *Hippocampus, 20*(11), 1217–1227.

Nyberg, L., Habib, R., Mcintosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America, 97*(20), 11120–11124.

Okado, Y., & Stark, C. E. L. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory, 12*(1), 3–11.

Packard, P. A., Rodríguez-Fornells, A., Bunzeck, N., Nicolás, B., de Diego-Balaguer, R., & Fuentemilla, L. (2017). Semantic congruence accelerates the onset of the neural signals of successful memory encoding. *The Journal of Neuroscience, 37*(2), 291–301.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976–987.

Pidgeon, L. M., & Morcom, A. M. (2016). Cortical pattern separation and item-specific memory encoding. *Neuropsychologia, 85*, 256–271.

Putnam, A. L., Sungkhasettee, V. W., & Roediger III, H. L. (2017). When misinformation improves memory: The effects of recollecting change. *Psychological Science, 28*(1), 36–46.

Richter, F. R., Cooper, R., Bays, P. M., & Simons, J. S. (2016). Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *eLife, 5*, e18260.

Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic plasticity: How memory retrieval drives learning. *Trends in Cognitive Sciences, 23*(9), 726–742.

Schacter, D. L., Guerin, S. A., & Jacques, P. L. S. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences, 15*(10), 467–474.

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49*(1), 289–318.

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review, 18*(3), 455–468.

Sekeres, M. J., Bonasia, K., St-Laurent, M., Pishdadian, S., Winocur, G., Grady, C., & Moscovitch, M. (2016). Recovering and preventing loss of detailed memory: Differential rates of forgetting for detail types in episodic memory. *Learning & Memory, 23*(2), 72–82.

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: prediction error influences episodic memory reconsolidation. *Learning & Memory, 25*(8), 369–381.

Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. *Trends in Neurosciences, 42*(10), 727–739.

Slotnick, S. D., & Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nature Neuroscience, 7*(6), 664–672.

Sommer, T. (2017). The emergence of knowledge and how it supports the memory for novel related information. *Cerebral Cortex, 27*(3), 1906–1921.

Spalding, K. N., Jones, S. H., Duff, M. C., Tranel, D., & Warren, D. E. (2015). Investigating the neural correlates of schemas: Ventromedial prefrontal cortex is necessary for normal schematic influence on memory. *Journal of Neuroscience, 35*(47), 15746–15751.

Staresina, B. P., Henson, R. N. A., Nikolaus, K., & Arjen, A. (2012). Episodic reinstatement in the medial temporal lobe. *Journal of Neuroscience, 32*(50), 18150–18156.

Stevenson, R. F., Reagh, Z. M., Chun, A. P., Murray, E. A., & Yassa, M. A. (2020). Pattern separation and source memory engage distinct hippocampal and neocortical regions during retrieval. *Journal of Neuroscience, 40*(4), 843–851.

St-Laurent, M., Abdi, H., Bondad, A., & Buchsbaum, B. R. (2014). Memory reactivation in healthy aging: evidence of stimulus-specific dedifferentiation. *Journal of Neuroscience, 34*(12), 4175–4186.

Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions, 8*(1), 35–35.

Sweegers, C. C. G., Coleman, G. A., van Poppel, E. A. M., Cox, R., & Talamini, L. M. (2015). Mental schemas hamper memory storage of goal-irrelevant information. *Frontiers in Human Neuroscience, 9*, 629–629.

van Buuren, M., Kroes, M. C. W., Wagner, I. C., Genzel, L., Morris, R. G. M., & Fernandez, G. (2014). Initial investigation of the effects of an experimentally learned schema on spatial associative memory in humans. *Journal of Neuroscience, 34*(50), 16662–16670.

van den Honert, R. N., McCarthy, G., & Johnson, M. K. (2016). Reactivation during encoding supports the later discrimination of similar episodic memories. *Hippocampus, 26*(9), 1168–1178.

van der Linden, M., Berkers, R. M. W. J., Morris, R. G. M., & Fernández, G. (2017). Angular gyrus involvement at encoding and retrieval is associated with durable but less specific memories. *Journal of Neuroscience, 37*(39), 9474–9485.

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *The Journal of Neuroscience, 30*(47), 15888–15894.

Warren, D. E., Jones, S. H., Duff, M. C., & Tranel, D. (2014). False recall is reduced by damage to the ventromedial prefrontal cortex: implications for understanding the neural correlates of schematic memory. *Journal of Neuroscience, 34*(22), 7677–7682.

Webb, C. E., Turney, I. C., & Dennis, N. A. (2016). What's the gist? The influence of schemas on the neural correlates underlying true and false memories. *Neuropsychologia, 93*, 61–75.

Weinstein, Y., McDermott, K. B., & Chan, J. C. (2010). True and false memories in the DRM paradigm on a forced choice test. *Memory, 18*(4), 375–384.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences of the United States of America, 97*(20), 11125–11129.

Wing, E. A., Geib, B. R., Wang, W. C., Monge, Z., Davis, S. W., & Cabeza, R. (2020). Cortical overlap and cortical-hippocampal interactions predict subsequent true and false memory. *Journal of Neuroscience, 40*(9), 1920–1930.

Xiao, X. Q., Dong, Q., Gao, J. H., Men, W. W., Poldrack, R. A., & Xue, G. (2017). Transformed neural pattern reinstatement during episodic memory retrieval. *Journal of Neuroscience, 37*(11), 2986–2998.

Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. L. (2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus, 21*(9), 968–979.

Ye, Z. F., Zhu, B., Zhuang, L. P., Lu, Z. L., Chen, C. S., & Xue, G. (2016). Neural global pattern similarity underlies true and false memories. *Journal of Neuroscience, 36*(25), 6792–6802.

Zhu, B., Chen, C. S., Shao, X. H., Liu, W. Z., Ye, Z. F., Zhuang, L. P., ... Xue, G. (2019). Multiple interactive memory representations underlie the induction of false memory. *Proceedings of the National Academy of Sciences of the United States of America, 116*(9), 3466–3475.

The cognitive and neural mechanisms underlying false memory: An information processing perspective

GUO Ying[1], GONG Xianmin[2], WANG Dahua[1]

([1] *Institute of Developmental Psychology, Beijing Normal University, Beijing* 100875, *China*)

([2] *Stanley Ho Big Data Decision Analytics Research Centre & Department of Psychology, The Chinese University of Hong Kong, Hong Kong* 999077, *China*)

Abstract: By reviewing both behavioral and neuroimaging research, the present article illustrates how processing of information from different sources (i.e., the target event/stimulus, internal cognitive schemas, and external interference) and at different stages (i.e., the encoding, storage, re-activation/reconsolidation and retrieval stages) contributes to false memory. We conclude that false memory may arise from three mechanisms: (1) The lack of distinctive item-specific memory representations that makes it difficult to distinguish targets from related lures; (2) The engagement of cognitive schemas strengthens the memory representations of non-target information (including related lures) in the schemas; and (3) Re-activated memory representations of targets are distorted and modified by external interference. Future research may use updated approaches, e.g., multivariate pattern analysis (MVPA), to further investigate the brain regions responsible for representing item-specific details, the way different types of schema (e.g., event-based script) promote the representations of related lures, and the way re-activation of schema during memory retrieval influences false memory.

Key words: false memory, cognitive schema, neural mechanisms, information processing